

Санкт-Петербургский государственный университет

СМАЧЕВ Кирилл Максимович

Выпускная квалификационная работа

**Методы интеллектуального анализа данных аварийных
событий и технологических нарушений в электрических
системах**

Направление 01.04.02

«Прикладная математика и информатика»

Основная образовательная программа ВМ.5504 «Исследование операций и
системный анализ»

Научный руководитель:
заведующий кафедрой
технологии
программирования,
к.т.н., доцент
Блеканов Иван
Станиславович

Рецензент: профессор
кафедры ЮНЕСКО по
ИВТ, ФГБОУ ВО КемГУ,
д.т.н., Степанов Юрий
Александрович

Санкт-Петербург
2020

Оглавление

Введение.....	3
Постановка задачи.....	5
Обзор литературы.....	7
Глава 1. Обзор решений.....	12
1.1. Обзор технических решений	12
1.2. Обзор методов	14
Глава 2. Разработка программного комплекса.....	15
2.1. Архитектура решения.....	15
2.1.1. Стек технологий	15
2.1.2. Структура модулей.....	20
2.2. Разработка аналитического компонента системы.....	21
2.3. Аprobация и тестирование	28
2.3.1. Постановка эксперимента	28
2.3.2. Результаты эксперимента	33
Выводы	43
Заключение	44
Результаты работы.....	44
Перспективы развития.....	45
Список литературы	47

Введение

Электроэнергетика является одной из важнейших областей хозяйственно-экономической деятельности человечества. Стабильность электроэнергетических систем, таких как электросети, в значительной степени влияет на качество большинства социальных и экономических процессов. Ключевым показателем стабильности электросети, является частота отказов. Даже кратковременные отказы электросети влекут за собой крупные экономические издержки и могут приводить, к чрезвычайным ситуациям и катастрофам. В связи с этим, прогнозирование отказов электросети является важной и по сей день актуальной задачей.

Большой вклад в развитие данной и смежных областей внес институт IEEE. Институт инженеров электротехники и электроники (Institute of Electrical and Electronics Engineers) — международная некоммерческая ассоциация специалистов в области техники, мировой лидер в области разработки стандартов по радиоэлектронике, электротехнике и аппаратному обеспечению вычислительных систем и сетей. Благодаря базе знаний IEEE Xplore существует большое количество публикаций на тему прогнозирования аварий, электротехнических и аналитических методов обнаружения неполадок в электросистемах, некоторые из которых были упомянуты в данной работе [12], [16], [9], [6], [26], [22], [23], [18]. IEEE Xplore — это исследовательская база данных для поиска и доступа к журнальным статьям, материалам конференций, техническим стандартам и связанным материалам по информатике, электротехнике и электронике и смежным областям.

Объектом исследования в данной работе является набор данных о топологической структуре европейской электросети, содержащий данные о расположении и связи станций и подстанций. А также данные, полученные с устройств автоматической релейной защиты и датчиков газоанализаторов, предоставляющих информацию о состоянии линий электропередач и трансформаторов. Данные являются многомерным временным рядом,

отображающим состояние системы в момент аварийных событий на подстанции и представленным в виде набора осциллограмм, в формате COMTRADE. Данная работа была реализована в рамках сотрудничества с отделом исследования и разработки, компании Siemens.

Постановка задачи

Целью работы является разработка системы для прогнозирования отказа компонентов электросети, на основе данных аварийных событий и технологических нарушений.

Для достижения данной цели необходимо выполнить ряд задач:

1. Разработка программного модуля обработки временных данных от компонентов электросети.
2. Построение алгоритма оценки компонентов электросети.
3. Построение алгоритма прогнозирования отказа компонента.
4. Разработка программного модуля визуализации электрокомпонентов на географической карте.

Для обработки исходных данных, поступающих с датчиков, отображающих состояние компонента, необходимо выполнить следующие задачи:

1. Разработка функций чтения осциллограмм в формате COMTRADE и сохранение в базе данных временных рядов Influx.
2. Разработка функций визуализации осциллограмм.
3. Разработка функций анализа данных об аварийных событиях и технологических нарушениях в электрических системах.

Для оценки наиболее важных узлов электросети необходимо выполнить следующие задачи:

1. Анализ топологической структуры электросети, построение ориентированного графа.
2. Определение свойств сети. Определение связности и центральности вершин.
3. Определение важности компонента сети, с точки зрения последствий его отключения от сети.
4. Оценка последствий селективного и случайного отключения компонентов.

Для построения алгоритма прогнозирования необходимо выполнить следующие задачи:

1. Сбор информации о физических свойствах компонентов электросети.
2. Обзор методов анализа данных об аварийных событиях и технологических нарушениях в электрических системах.
3. Обоснование выбора метода прогнозирования.
4. Реализация алгоритма обработки исходных данных.
5. Реализация алгоритма прогнозирования.

Обзор литературы

Терминология, используемая в данной работе, опирается на государственные стандарты в данной области [2], [3], [4], [5]. В соответствии с данными стандартами, в работе употребляется ряд определений.

Электростанция (ЭС) – энергоустановка, предназначенная для производства электрической энергии, содержащая строительную часть, оборудование для преобразования энергии и необходимое вспомогательное оборудование по ГОСТ 19431 [3].

Линия электропередачи (ЛЭП) – электроустановка, состоящая из проводов, кабелей, изолирующих элементов и несущих конструкций, предназначенная для передачи электрической энергии между двумя пунктами энергосистемы с возможным промежуточным отбором по ГОСТ 19431 [3].

Распределительное устройство (РУ) – электроустановка, предназначенная для приема и распределения электрической энергии на одном напряжении и содержащая коммутационные аппараты и соединяющие их сборные шины, устройства управления и защиты.

Примечание. К устройствам управления относятся аппараты и связывающие их элементы, обеспечивающие контроль, измерение, сигнализацию и выполнение команд [3].

Подстанция (ПС) – электроустановка, предназначенная для приема, преобразования и распределения электрической энергии, состоящая из трансформаторов или других преобразователей электрической энергии, устройств управления, распределительных и вспомогательных устройств по ГОСТ 19431 [3].

Электропередача – совокупность линий электропередачи и подстанций, предназначенная для передачи электрической энергии из одного района энергосистемы в другой [3].

Электрическая сеть – совокупность подстанций, распределительных устройств и соединяющих их линий электропередачи, предназначенная для передачи и распределения электрической энергии по ГОСТ 19431 [3].

Релейная защита (РЗ) – совокупность устройств, предназначенных для автоматического выявления коротких замыканий, замыканий на землю и других ненормальных режимов работы ЛЭП и оборудования, которые могут привести к их повреждению и/или нарушению устойчивости энергосистемы, формирования управляющих воздействий на отключение коммутационных аппаратов в целях отключения этих ЛЭП и оборудования от энергосистемы, формирования предупредительных сигналов [4].

Релейная защита и автоматика (РЗА) – релейная защита, сетевая автоматика, противоаварийная автоматика, режимная автоматика, регистраторы аварийных событий и процессов, технологическая автоматика объектов электроэнергетики [4].

Регистраторы аварийных событий и процессов – устройства, регистрирующие аварийные события и процессы в энергосистеме (регистраторы аварийных событий, регистраторы системы мониторинга переходных процессов, устройства определения места повреждения) [4].

Трансформатор – статическое электромагнитное устройство, имеющее две или более индуктивно связанных обмоток и предназначенное для преобразования посредством электромагнитной индукции одной или нескольких систем переменного тока в одну или несколько других систем переменного тока [5].

Под узлами электросети будут подразумеваться электростанции и электрические подстанции, предназначенные для передачи сигнала по линиям электропередач, основным элементом которых, являются трансформаторы тока и напряжения. Под связями электросети подразумеваются линии электропередач.

Существует общепринятый стандарт записи аварий в электросистемах, разработанный ассоциацией IEEE (Institute of Electrical and Electronics

Engineers). Этот стандарт определяет общий формат для файлов данных и средств передачи, необходимых для обмена различными типами данных повреждения, тестов и моделирования [12]. Обработка данных аварий не была бы возможна без информации о принятых форматах записей осциллограмм.

Отправной точкой для исследования вопроса оценки важности электрокомпонента в электросети стала статья «Robustness of the European power grids under intentional attack» [21]. В данной статье был продемонстрирован подход к анализу атак электросети. Под атакой подразумевается множественное удаление узлов и последующая оценка состояния сети. Такой подход моделирует реакцию электросети в следствии случайного или умышленного повреждения. Однако, в данной работе не рассматривался вопрос об оценке важности узла сети, с точки зрения его отключения, а лишь демонстрировалось общее состояние системы, вследствие атак. В главе 2 будет показан новый метод оценки важности узла сети.

В учебнике Алексеева В.Е. и Таланова В.А. «Графы. Модели вычислений. Структуры данных» хорошо изложены фундаментальные понятия теории графов, которые необходимы при исследовании топологической структуры электросети [1].

Современные алгоритмы для анализа графов, и, в частности, расчета метрик центральности, используемые в сравнении, изложены в книге Matthew O. Jackson. «Social and Economic Networks» [16].

Примером практической реализации графа в графовой СУБД, является статья «Topology Modeling and Analysis of a Power Grid Network Using a Graph Database» [9]. Данная работа демонстрирует общие принципы работы графовых СУБД.

Способ детектирования аварий трансформаторов, основанный на анализе растворённых газов, а также физические явления, связанные с отказом трансформаторов, описываются в работе «Power Transformer Condition Monitoring and Diagnosis» [6].

Анализ газов, растворенных в масле, является экспертным методом, служащим для обнаружения дефектов и аварий узлов электрооборудования.

Для оценки состояния оборудования используются следующие газы: водород (H_2), метан (CH_4), этан (C_2H_6), этилен (C_2H_4), ацетилен (C_2H_2), угарный газ (CO), углекислый газ (CO_2). Вне зависимости от состояния оборудования, наблюдается наличие кислорода и азота, концентрация данных газов может изменяться в зависимости от реализации трансформатора. Экспертами, степень поврежденности трансформатора, оценивается по средства сопоставления полученных при анализе данных с референсными значениями концентрации газов в масле.

Данная информация позволяет сделать вывод о том, что имеет смысл рассмотреть применение методов классификации для определения вероятности аварии в трансформаторе.

В статье «Distribution Transformer Health Monitoring and Predictive Asset Maintenance» показан пример системы, прогнозирующей возникновение отказа трансформатора. Особенностью данной системы является использование данных о предыдущих авариях трансформатора, простоях трансформатора и погоды [19]. Подход, показанный в главе 2, отличается тем, что базируется на других данных, дает вероятностную оценку возникновения аварии, а также возможностью дать оценку узлу сети, без использования исторической информации.

В статье «A Transformer Fault Diagnosis Method Based on Bayesian Network» рассматривается подход к анализу аварий трансформаторов, с использованием байесовских сетей [26].

В работах «Transformer Reliability Taking Predictive Maintenance Program to the Next Level», «Condition assessment of power transformers», «Failure Modes and Effects Analysis (FMEA) for Power Transformers» и «Failure analysis of power transformers for effective maintenance planning in electric utilities» рассматривается использование электротехнических методов для

диагностирования и прогнозирования аварий трансформаторов [12], [25], [17], [20].

Существуют публикации на тему детектирования аварий на линиях электропередач. Электротехнические методы рассматриваются в статьях «Transmission line fault detection and location using Wide Area Measurements» и «Transmission Line Fault Detection: A Review» [22], [14].

Методы машинного обучения показали свою эффективность в детектировании аварий на линиях электропередач в работах «A real-time fault detection and classification algorithm for transmission line faults based on MODWT during power swing» и «Fault detection, location and classification of a transmission line» [24], [20]. Новый подход, описанный в главе 2, отличается использованием статистических параметров сигнала, в качестве исходных данных для классификации.

Глава 1. Обзор решений

1.1. Обзор технических решений

Существуют решения, позволяющие проводить анализ топографической структуры электросети, использованием графовых СУБД. В частности, Neo4j [9].

Neo4j17, открытая графическая база данных с открытым исходным кодом, написанная на Java и Scala, реализует общие графические модели с полными характеристиками базы данных, такими как соответствие транзакциям ACID, поддержка кластеров, обход отказа во время выполнения и язык запросов. Она обозначает различные роли по узлам с множественными метками и представляет отношения по путям между узлами.

Данное решение было сравнено с классическими реляционными подходами, было показано, что графические базы данных могут достичь значительного повышения производительности для сложных отношений данных. Была построена графовая структура базы данных для электросетевой сети на базе Neo4j, а также разработана система для анализа первичной энергообеспеченности электросети. Эксперименты показали, что анализ энергосетей с использованием Neo4j имеет лучшую производительность, чем с использованием PostgreSQL. Кроме того, в данном решении реализована функция поиска кратчайшего пути на основе Neo4j's Cypher и функция условного поиска на основе Neo4j's Traversal API.

По части детектирования аварий в электросистемах на основе методов машинного обучения, существует реализации на основе таких решений как SAS Enterprise Guide, SAS Enterprise Miner и SAS Visual Analytics [19]. Методология, представленная в данном решении, обнаруживает и визуализирует отключения и перегруженные участки в режиме, близком к реальному времени обнаруживает трансформаторы и распределительные участки с длительной историей отключений и перегрузок. Данное решение

реализует прогнозные алгоритмы для улучшения регулярных графиков технического обслуживания оборудования и сокращения количества поездок ремонтных машин на внеплановое техническое обслуживание во время незапланированных инцидентов, таких как ураганы.

Отличием, предложенного в главе 2 решения от вышеперечисленных, является одновременная реализация трех наиболее важных задач:

1. Оценка важности узла сети.
2. Прогноз аварии в узле сети.
3. Прогноза аварии в связях, между узлами.

1.2. Обзор методов

Как показал обзор существующих научных работ в области, на данный момент, существуют, как автоматические методы определения неисправностей в энергосистемах, так и менее формальные методы, используемые экспертами. В данной работе пойдёт речь об алгоритмах автоматического детектирования аварийной ситуации:

1. метод опорных векторов [14], [23], [15], [18];
2. логистическая регрессия [10], [18],[27];
3. наивный байесовский классификатор [26];
4. метод К-ближайших соседей [14], [7], [8];
5. дерево решений [19].

Методы на основе байесовских сетей и деревьев решений продемонстрировали свою эффективность задачах детектирования аварий [19], [26]. В качестве дополнительных методов, для классификации аварий рассматриваются логистическая регрессия, метод К-ближайших соседей и метод опорных векторов, [14], [7], [8], [23], [15], [18]. Надо заметить, что все перечисленные методы являются классическими для задач, связанных с классификацией, и зарекомендовали себя в качестве эффективных методов, для решения подобных задач [19], [26], [14], [7], [8], [23], [15], [18].

Глава 2. Разработка программного комплекса

2.1. Архитектура решения

2.1.1. Стек технологий

Технологией, применяемой для записи и хранения состояний электросистемы во время аварии, является формат COMTRADE. Поскольку информация о технологических нарушениях может поступать от разнообразных источников, различных производителей, использующих собственные или другие стандартные форматы, общий формат необходим для облегчения обмена такими данными между устройствами с различными прикладными системами, которые могут использовать цифровые данные от других устройств.

Данный формат представляется тремя файлами. Файл заголовка создается источником данных аварийного процесса с помощью программ текстовой обработки. Файл конфигурации предназначен для чтения компьютерной программой и, следовательно, должен иметь информацию в определенном формате. Данные в файле конфигурации – это информация, необходимая компьютерной программе для правильной интерпретации данных переходного процесса. Файл данных содержит значение каждой выборки каждого входного канала. Число, записанное как выборка обычно число, выданное устройством, которое производит выборки входного переменного сигнала.

Пример диаграммы, записанной в формате COMTRADE представлен на рисунке 1,2.

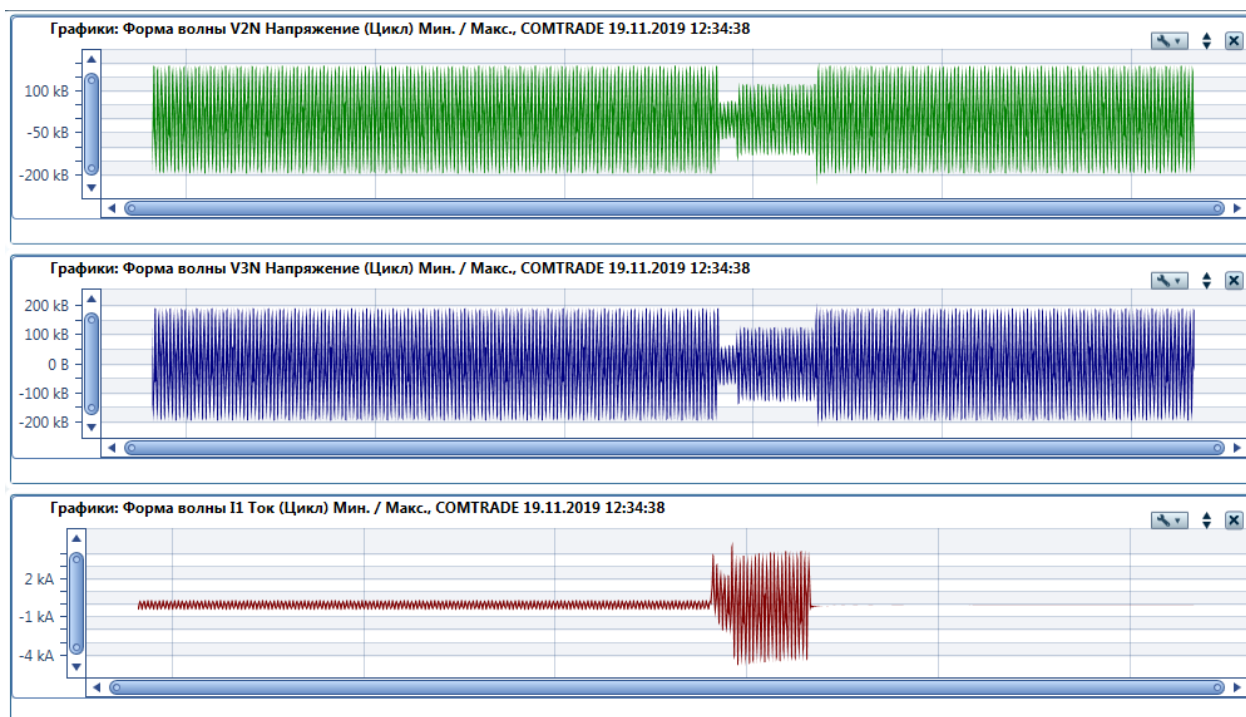


Рисунок 1 – Пример диаграммы в формате COMTRADE, аналоговый сигнал

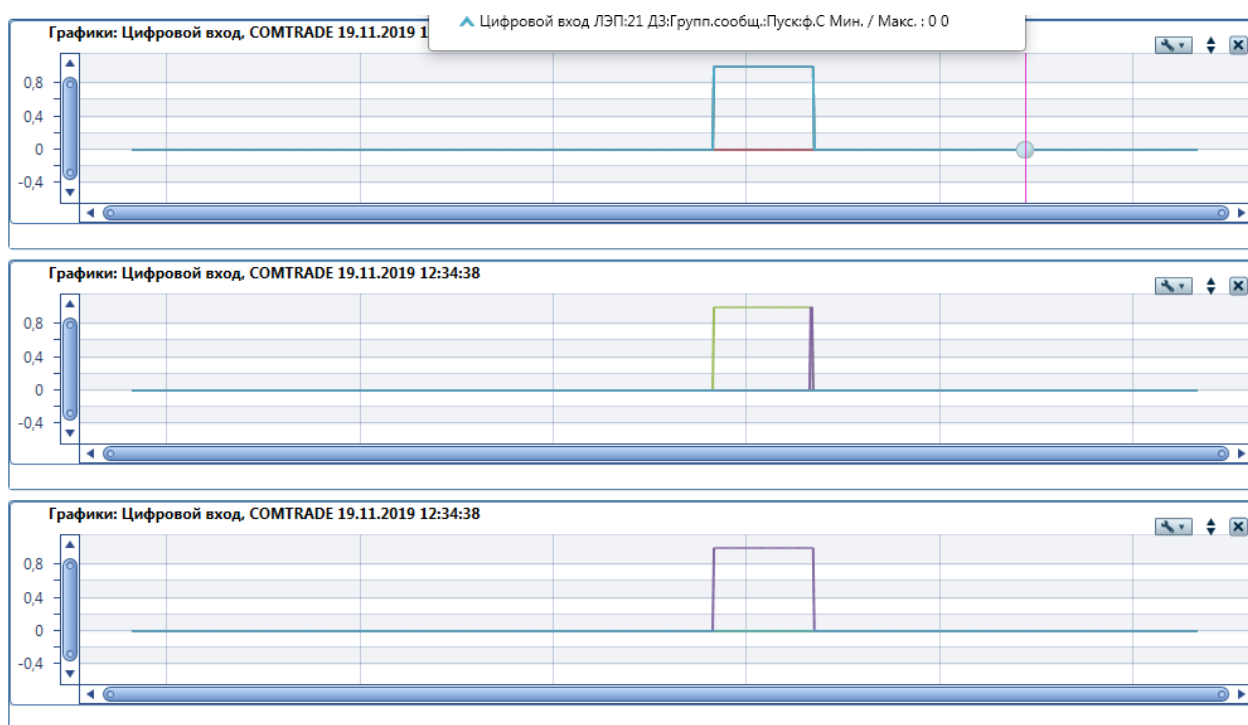


Рисунок 2 – Пример диаграммы в формате COMTRADE, цифровой сигнал

Реализация сбора данных была произведена посредством языка программирования Python, базы данных временных рядов InfluxDB и инструмента визуализации Grafana.

InfluxDB представляет собой программное обеспечение с открытым исходным кодом для хранения временных рядов. Основным назначением InfluxDB является хранение больших объемов данных с метками времени.

Grafana – веб-интерфейс для работы с графиками. Позволяет визуализировать данные с систем мониторинга и метрик. Пример отображения исходных данных представлен на рисунке 3.

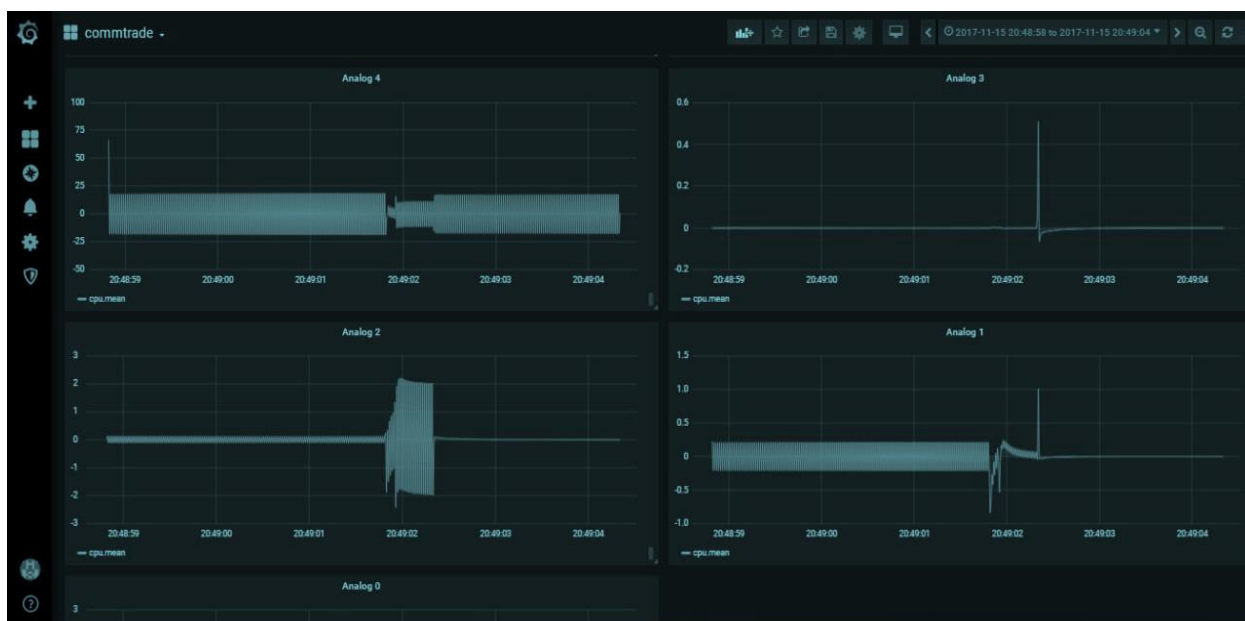


Рисунок 3 – Пример отображения исходных данных в Grafana

Было разработано приложение на языке Python, позволяющее осуществить синтаксический анализ и визуализацию диаграмм данных.

Данные поступают с устройств, установленных на компонентах электросети, таких, как трансформаторы и генераторы.

Для работы с графом использовалась графовая СУБД ArangoDB. Исходные данные хранятся в файле, формата *.csv. Экспорт данных осуществлялся посредством модуля экспорта, разработанного на языке Python.

ArangoDB – это распределенная СУБД, разработанная triAGENS GmbH. Система базы данных поддерживает три важные модели данных с одним ядром базы данных и единым языком запросов AQL (язык запросов

ArangoDB). Язык запросов является декларативным и позволяет сочетать разные шаблоны доступа к данным в одном запросе. Эта БД позволит создавать высокопроизводительные приложения, используя удобный язык запросов SQL или JavaScript.

На рисунках 4,5 представлен фрагмент, топологической структуры графа, визуализация средствами ArangoDB.

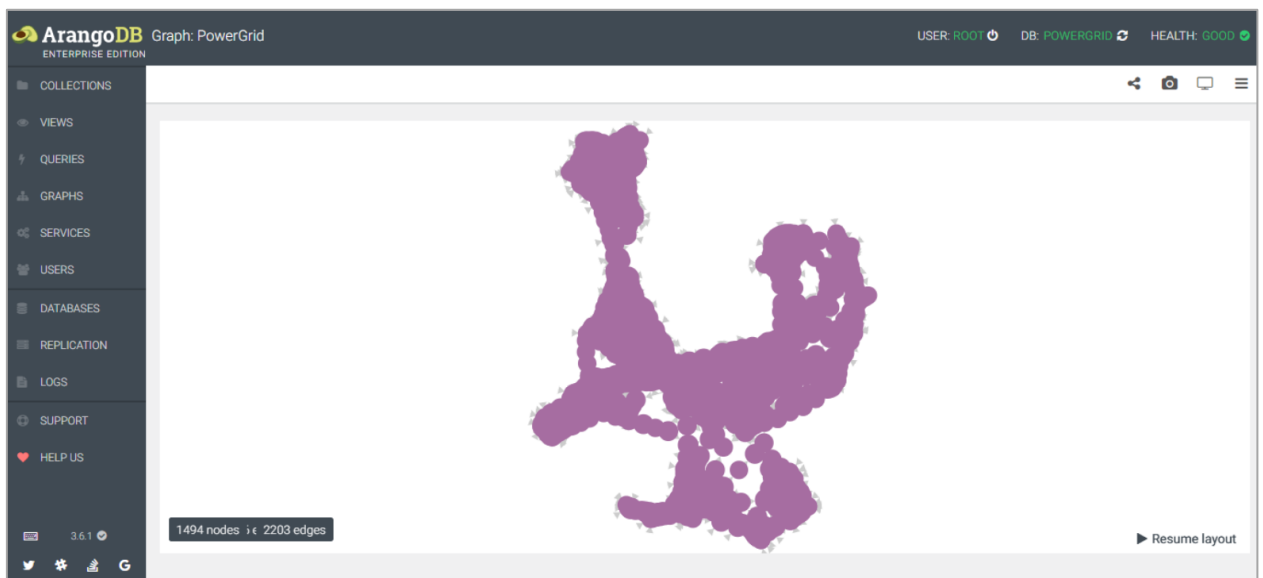


Рисунок 4 – Визуализация топологической структуры графа ArangoDB

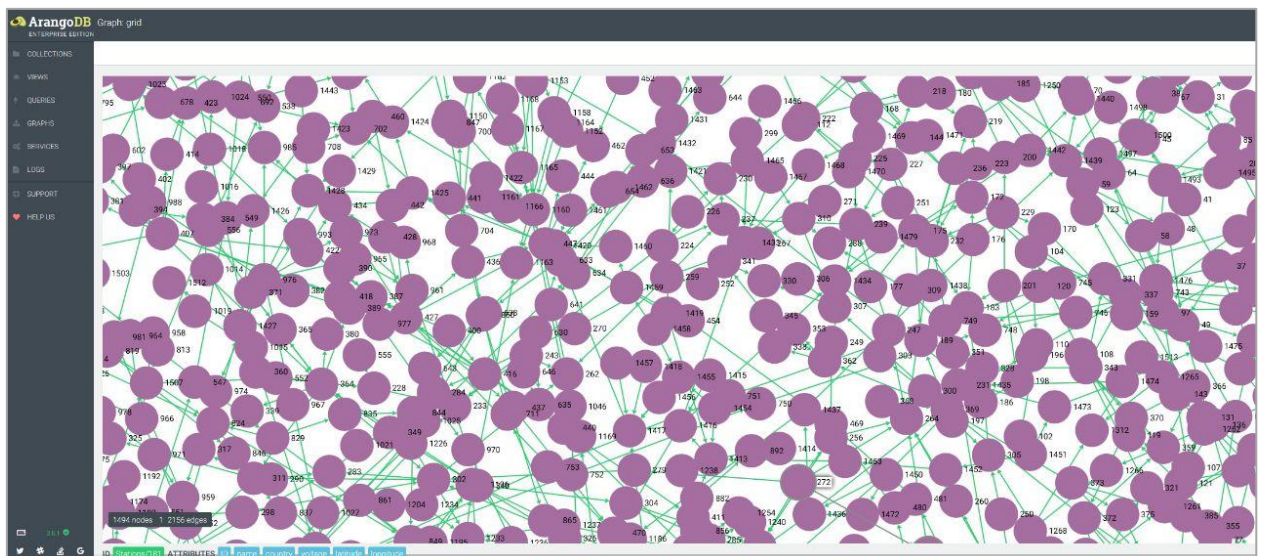


Рисунок 5 – Визуализация фрагмента топологической структуры графа ArangoDB

Данная СУБД использует формат JSON для работы с данными. Геоданные хранятся в формате Geojson. Географическая структура графа представлена на рисунках 6,7.

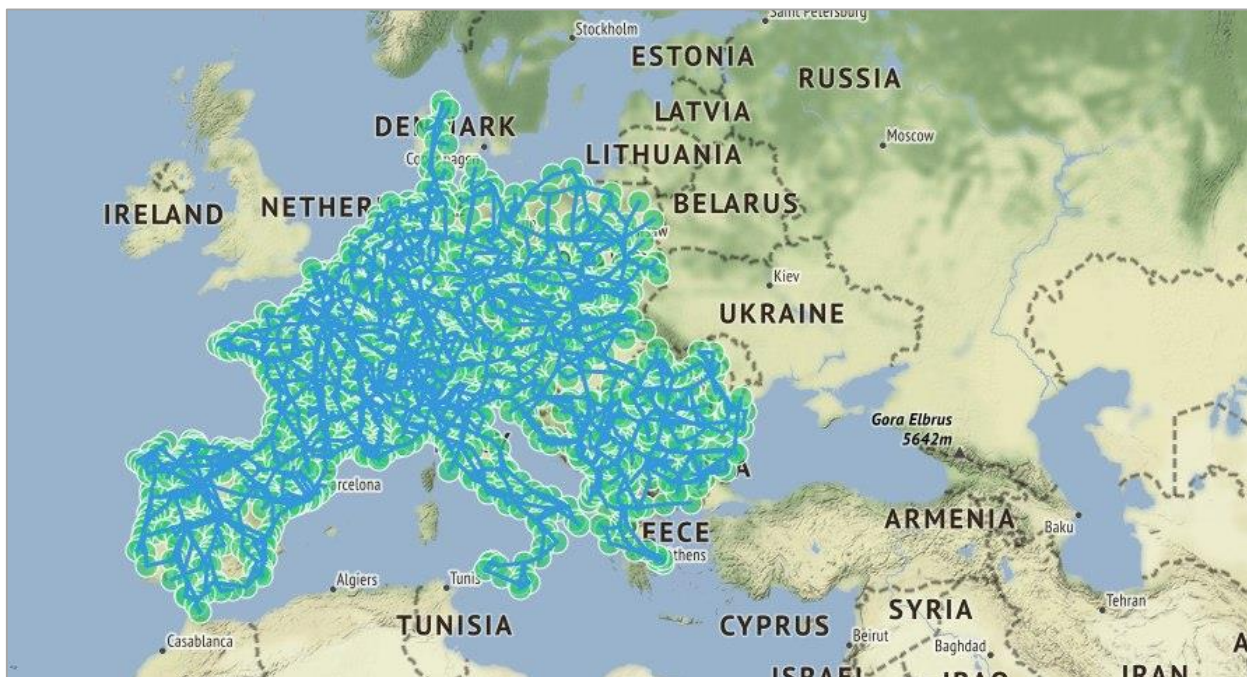


Рисунок 6 – Визуализация географической структуры графа ArangoDB

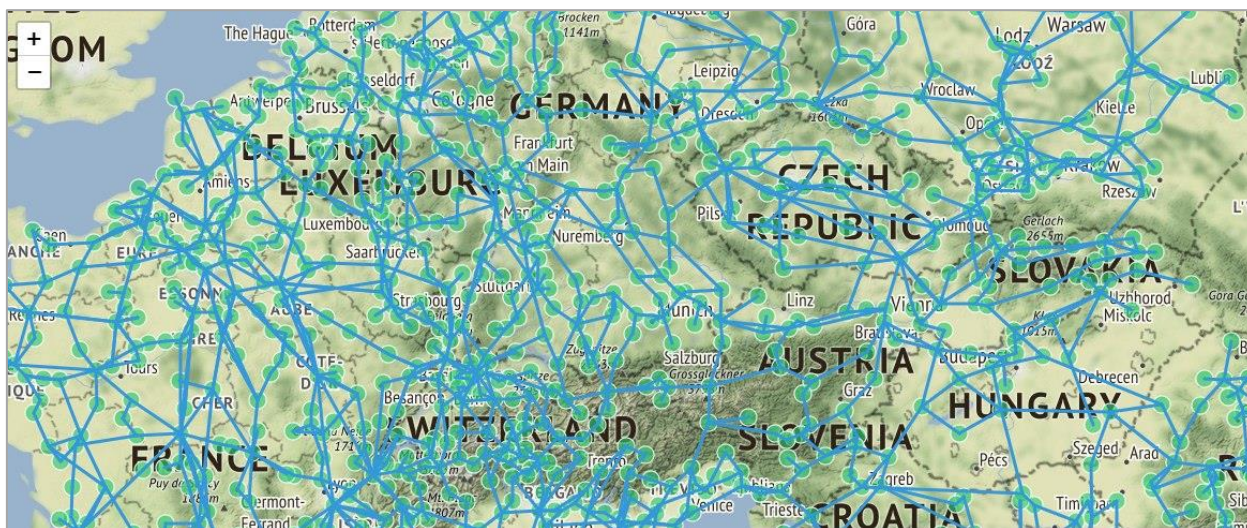


Рисунок 7 – Фрагмент визуализации географической структуры графа ArangoDB

2.1.2. Структура модулей

Таким образом, основным связующим звеном системы является разработанное приложение на языке, реализующее сбор, и распределение исходных данных по остальным системам.

В первую очередь данные о состоянии узлов и связей электросети обрабатываются и загружаются в базу данных InfluxDB. После чего визуализируются в система Grafana. Данные о структуре графа электросети поступают в базу данных ArangoDB и визуализируются посредством запросов AQL. После этого происходит расчет вероятности отказа для узлов и связей сети, с последующим выводом информации в ArangoDB.

Итоговая структура модулей разработанного программного обеспечения представлена на рисунке 8.

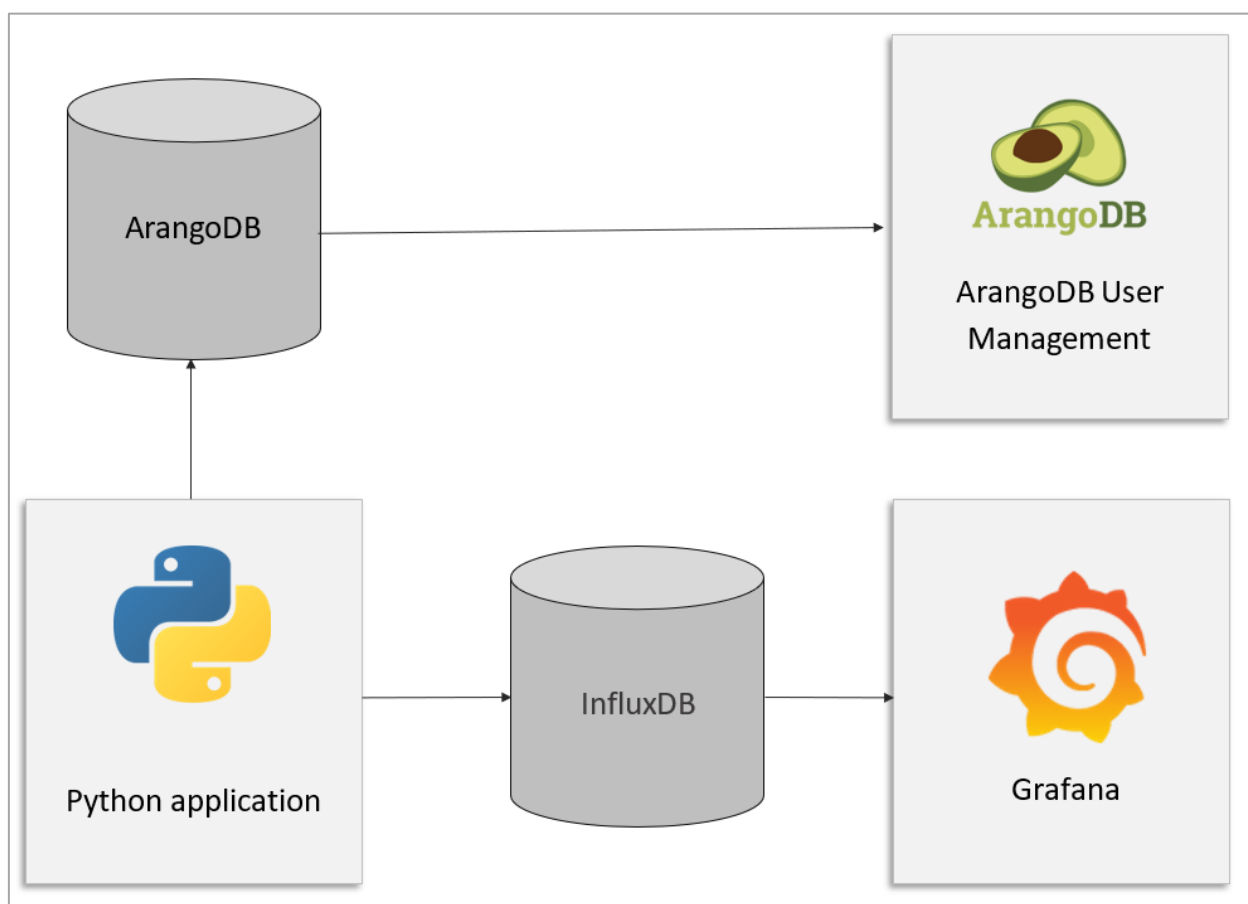


Рисунок 8 – Схема подключения модулей

2.2. Разработка аналитического компонента системы

Для разработки метода оценки важности узла, с точки отключения от сети, рассмотрим граф, отображающий топологическую и географическую структуру электросети. Вершинами графа являются электростанции, то есть генераторы электрического сигнала, и подстанции, элементы, необходимые для распределения электричества. Ребра графа представляют собой линии передачи электроэнергии такие, как ЛЭП и т.д. Данный граф ориентированный, однако для применение нижеперечисленных алгоритмов была построена его неориентированная копия. Полученный граф состоит из одной компоненты связности.

Плотный граф – граф, в котором число рёбер близко к максимально возможному у полного графа с числом V вершин. Граф, имеющий малое число рёбер, принято называть разреженным графом [1].

Для неориентированного простого графа (реберная) плотность графа с числом вершин V определяется как отношение числа его рёбер E к числу ребер полного графа, формула 1:

$$D = \frac{2E}{V(V-1)} \quad (1)$$

Максимальное число ребер определяется по формуле 2:

$$E = \frac{V(V-1)}{2} \quad (2)$$

Так что максимальная плотность графа равна 1 (для полных графов) и минимальная равна 0 – для несвязного графа.

Сообщества – подграфы исходного графа, имеющие наибольшую плотность. При помощи алгоритма Label Propagation, граф разбивается на n сообществ.

Также база хранит информацию о странах, в которых располагаются компоненты, вершины с группировкой по странам пересекаются, но не совпадают с множеством сообществ.

Самые крупные сообщества в порядке убывания размера представлены на рисунках 9-15.



Рисунок 9 – Сообщество 1 размер 161



Рисунок 10 – Сообщество 2 размер 152



Рисунок 11 – Сообщество 3 размер 151

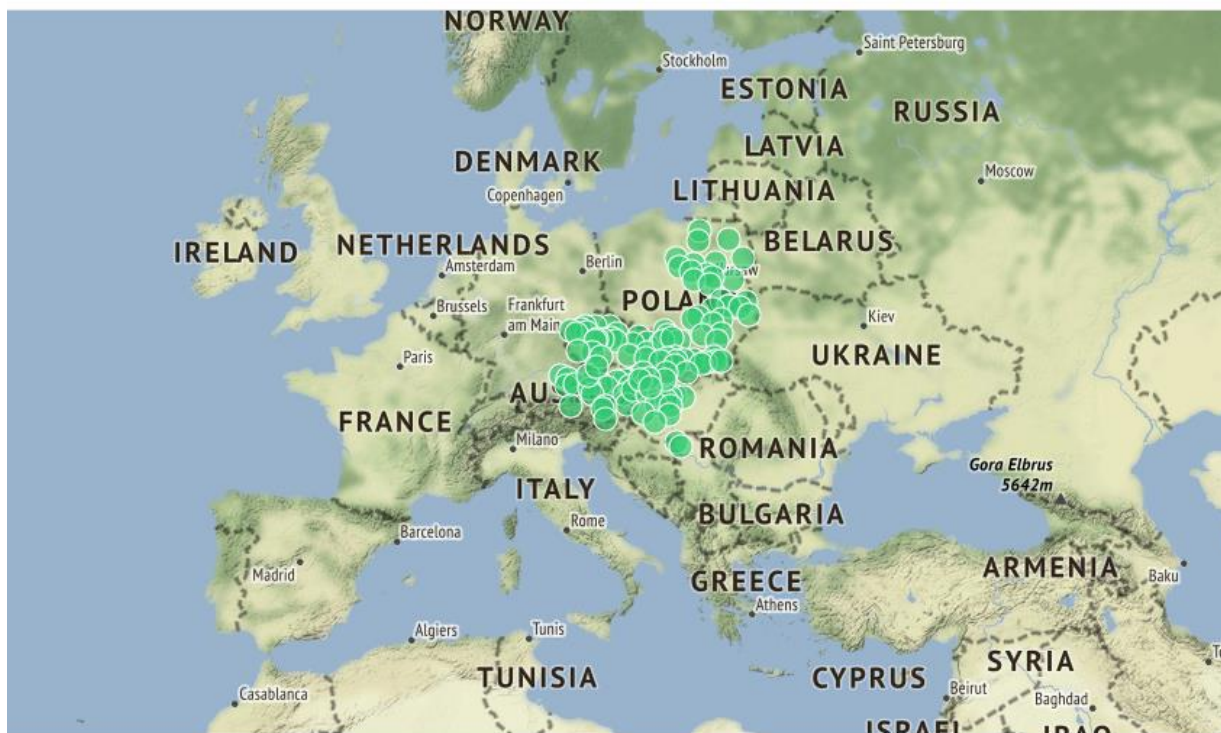


Рисунок 12 – Сообщество 4 размер 112

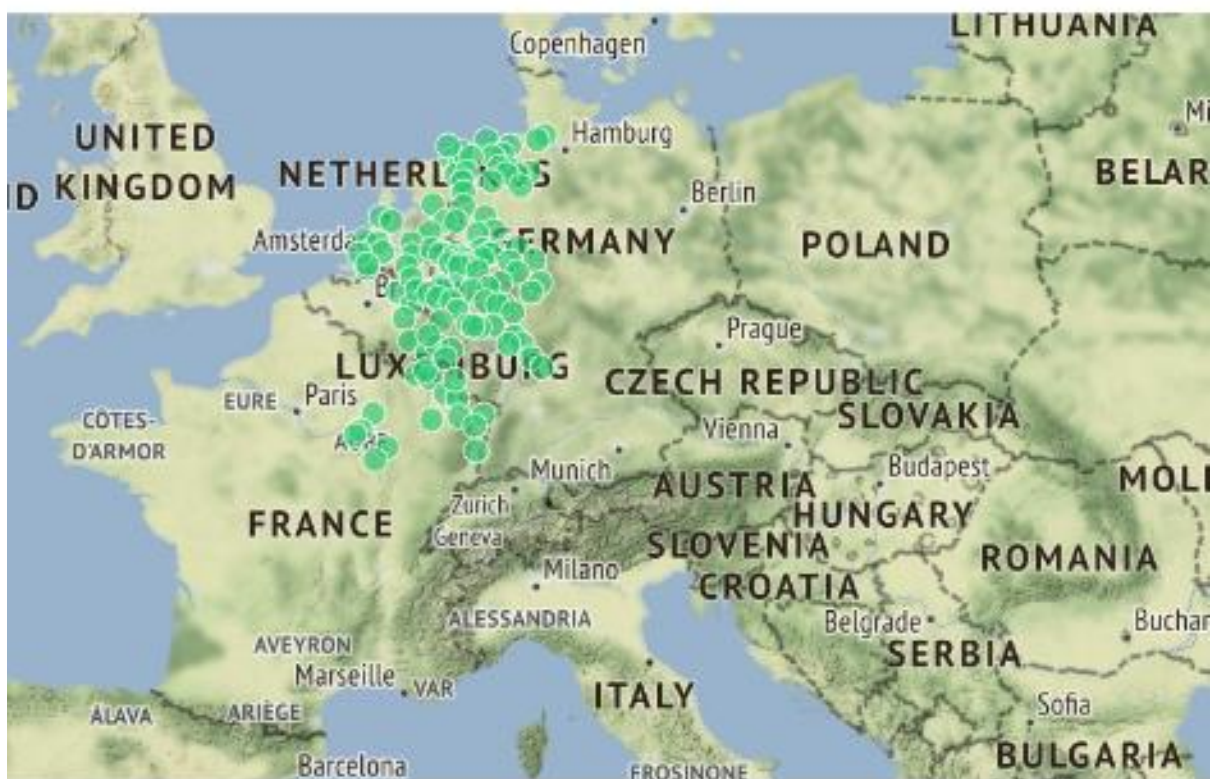


Рисунок 13 – Сообщество 5 размер 110



Рисунок 14 – Сообщество 6 размер 100



Рисунок 15 – Сообщество 7 размер 48

Разумно предположить, что часть графа, не причастная к большим сообществам, содержит вершины, удаление которых может быть критично для

связи между сообществами и нанести наибольший ущерб. Однако все предыдущие шаги, подразумевают использование уже известных алгоритмов.

Разработан новый метод, для определения важности вершины, данный метод основан на приросте количества компонент связности.

Алгоритм разработанного метода представлен ниже:

1. Для $j = 1..K$
2. Для $i = 1..N$,
3. Отключить вершину i от сети,
4. Рассчитать количество компонент связности C_i , после отключение вершины i ,
5. Рассчитать прирост количества компонент связности, после отключения вершины dC_i ,
6. Подсчитать среднее dC_i по j .

Где K – количество итераций, задается заранее.

N – количество вершин.

C_i – количество компонент связности после отключения вершины i .

dC_i – прирост компонент связности после отключения вершины i .

Рациональность применения данного алгоритма описана в параграфе «Результаты эксперимента». Недостатком данного алгоритма является вычислительное время, поскольку для установления точного результата необходимо перебрать все комбинации вершин, то есть $M = n!$.

Методы детектирования аварий на линиях электропередач и в трансформаторах будут основаны на нижеперечисленных методах машинного обучения:

1. метод опорных векторов;
2. логистическая регрессия;
3. наивный байесовский классификатор;
4. метод K -ближайших соседей;
5. дерево решений.

Новшеством в данной системе является способ обработки исходных данных. Для детектирования аварий на линиях электропередач, будут использоваться рассчитанные системой параметры сигнала. Посредством метода медианной фильтрации отделен шум и разобраны его статистические параметры. Такие как дисперсия, среднее, максимальное и минимальное значение выборки. В параграфах «Постановка эксперимента» и «Результаты эксперимента» представлено обоснование рациональности применения данной методологии.

2.3. Апробация и тестирование

2.3.1. Постановка эксперимента

Исходные данные не содержат электротехнических параметров, таких как, время восстановления системы после отключения компонента. Также нет информации о связях графа, например пропускная способность линии электропередач.

Это означает, что нет возможности использовать технические данные системы, а также алгоритмы, основывающиеся на весе ребер.

Исходя из этого в качестве метрики для оценки ущерба было выбрано количество компонент связности.

Увеличение компоненты связности говорит о том, что часть сети была отключена. Это влечет за собой пропорциональный экономический ущерб.

Поэтому для определения важности каждой вершины, были подсчитаны следующие параметры:

1. степень близости degree;
2. степень близости closeness.

Степень близости узла – это мера центральности в сети, вычисляемая как обратная величина суммы длин кратчайших путей между узлом и всеми другими узлами графа. Таким образом, чем более централен узел, тем ближе он ко всем другим узлам.

Количество компонент неориентированного исходного графа равно единице.

Для оценки последствий селективного и случайного отключения компонентов был применен следующий алгоритм:

1. Для $i = 1..n$,
2. Отключить случайную вершину i от сети,
3. Рассчитать количество компонент связности C_i , после отключение вершины i .

Результат выполнения данного алгоритма будет варьироваться, в зависимости от порядка отключения вершин.

В случае случайного отключения этот порядок определяется случайным образом. Данная ситуация моделирует стихийные события, которые могут повлиять на состояние электросети.

Однако, если опираться на такие параметры, как степень близости вершины, можно оценить, насколько каждый из компонентов сети важен, с точки зрения последствий его отключения. Тогда алгоритм отключения будет следующий:

1. Для $i = 1..n$,
2. Найти наиболее центральную вершину i , используя degree или closeness центральности.
3. Отключить вершину i от сети,
4. Рассчитать количество компонент связности C_i , после отключение вершины i .

Данный алгоритм, позволяет нанести гораздо больший ущерб сети, это может моделировать поведение злоумышленника.

Также, определив самые важные, с точки зрения последствий отключения, компоненты сети можно провести электротехнические работы по укреплению данных узлов, что поможет снизить ущерб от отключения.

Разработан новый метод, для определения важности вершины, данных метод основан на приросте количества компонент связности. Алгоритм разработанного метода представлен ниже:

1. Для $j = 1..K$
2. Для $i = 1..N$,
3. Отключить вершину i от сети,
4. Рассчитать количество компонент связности C_i , после отключение вершины i ,
5. Рассчитать прирост количества компонент связности, после отключения вершины dC_i ,

6. Подсчитать среднее dC_i по j .

Где K это количество итераций, задается заранее.

N – количество вершин.

C_i – количество компонент связности после отключения вершины i .

dC_i – прирост компонент связности после отключения вершины i .

Недостатком данного алгоритма является вычислительное время, поскольку для установления точного результата необходимо перебрать все комбинации вершин, то есть $M = n!$.

Как и говорилось ранее, метод для детектирования аварий и определения вероятности отказа узла будет основан на данных анализа растворенных газов. Взаимосвязь основных газов и наиболее характерных видов дефектов представлена в таблице 1.

Таблица 1 – Взаимосвязь характерных видов дефектов с растворенными в трансформаторе газами

Водород (H_2)	Дефекты электрического характера: частичные разряды, искровые и дуговые разряды
Метан (CH_4)	Дефекты термического характера: нагрев масла и бумажно-масляной изоляции в диапазоне температур $(400-600)^\circ C$ или нагрев масла и бумажно-масляной изоляции, сопровождающийся разрядами
Этан (C_2H_6)	Дефекты термического характера: нагрев масла и бумажно-масляной изоляции в диапазоне температур $(300-400)^\circ C$
Этилен (C_2H_4)	Дефекты термического характера: нагрев масла и бумажно-масляной изоляции выше $600^\circ C$
Ацетилен (C_2H_2)	Дефекты электрического характера: электрическая дуга, искрение
Угарный газ (CO)	Дефекты термического характера: старение и увлажнение масла и/или твердой изоляции
Углекислый газ (CO_2)	Дефекты термического характера: <ul style="list-style-type: none">• старение и увлажнение масла и/или твердой изоляции;• нагрев твердой изоляции.

На данный момент существуют классификации неполадок, в компонентах электросети. Например, анализ растворенных газов, позволяет определить вероятности одной из следующих ошибок, исходя из содержания и концентрации того или иного газа в трансформаторе.

Виды неполадок:

1. частичный разряды (PD);
2. разряды низкой энергии (D1);
3. разряды высокой энергии (D2);
4. тепловая неисправность менее 300 °C (T1);
5. тепловая неисправность между 300 °C и 700 °C (T2);
6. тепловой разлом выше 700 °C (T3).

Исходные данные содержат информацию о количестве, растворенных газов в маслах трансформатора в момент аварии и в нормальном состоянии, газов:

1. водород (H₂);
2. метан (CH₄);
3. этан (C₂H₆);
4. этилен (C₂H₄);
5. ацетилен (C₂H₂).

Каждый элемент выборки (аварийный случай) содержит метку принадлежности к определенному классу аварии:

1. частичные разряд (partial discharge);
2. искровые разряды (spark discharge);
3. дуговые разряды (arc discharge);
4. высокотемпературный перегрев (high-temperature overheating);
5. среднетемпературный перегрев (middle-temperature overheating);
6. низкотемпературный перегрев (low-temperature overheating).

Таким образом, данные представляют собой информацию об авариях в трансформаторах и количестве того или иного газа, зарегистрированного при соответствующей аварии. Пример данных представлен в таблице 2.

Таблица 2 – Пример исходных данных

NM	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂	Type	Typeid
1	3930	2397	157	0	0	Partial discharge	0
2	37800	1740	249	8	8	Spark discharge	1
3	92600	10200	0	0	0	Arc discharge	2
4	8266	1061	22	0	0	Partial discharge	0
5	9340	995	60	6	7	High-temp. Ov-ting	3

Для классификации исходная выборка была случайным образом разбита на обучающую и тестовую. Размер обучающей выборки – 85% от общего количества измерений. Размер тестовой выборки – 15% от общего количества измерений.

В качестве метрики будем рассматривать F-меру. F-мера представляет собой гармоническое среднее между точностью и полнотой. Она стремится к нулю, если точность или полнота стремится к нулю. Где точность и полнота для каждого класса. Точность равняется отношению соответствующего диагонального элемента матрицы и суммы всей строки класса. Полнота – отношению диагонального элемента матрицы и суммы всего столбца класса.

Под связями сети подразумеваются линии электропередач. Линия электропередачи (ЛЭП) – один из компонентов электрической сети, система энергетического оборудования, предназначенная для передачи электроэнергии посредством электрического тока.

Исходные данные содержат записи токового сигнала о состоянии линии электропередач в момент аварии, с метками 0 и 1, показывающими наличие повреждения аварии.

Пример графика исходных данных, зависимости величины сигнала I от времени t представлен на рисунке 16.

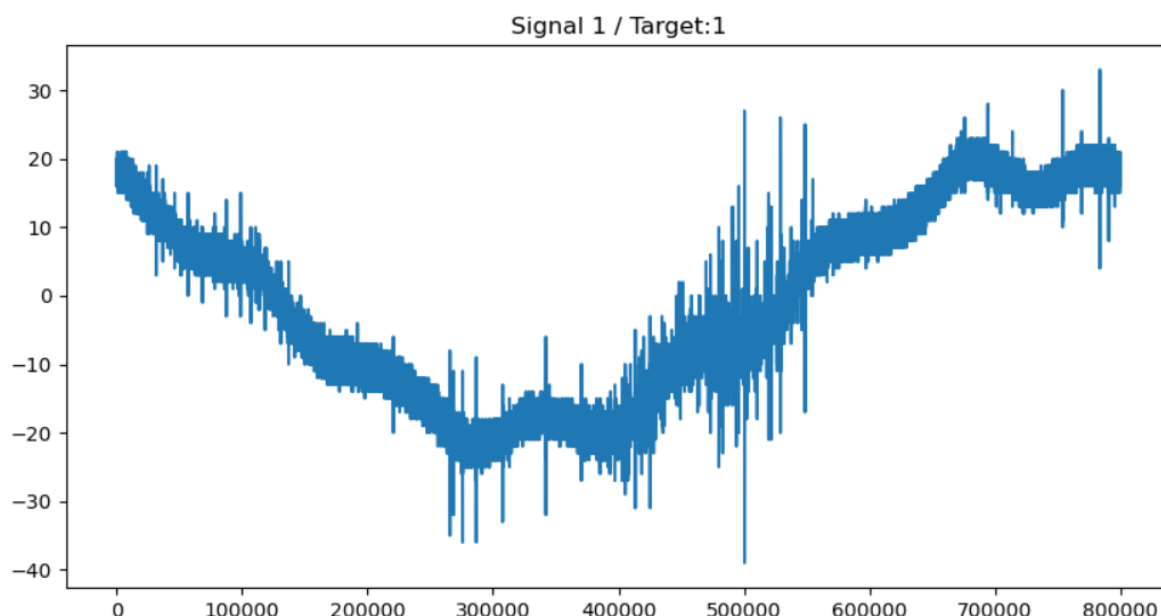


Рисунок 16 – Пример исходных данных

2.3.2. Результаты эксперимента

Результат сравнения разработанного алгоритма оценки важности с уже известными метриками, представлен на рисунке 17. Можно наблюдать, что отключение узлов, имеющих наибольшую степень важности, исходя из разработанного алгоритма ведёт к наиболее быстрому росту количества компонент связности. На практике это означает, что наибольший урон электросети принесет отключение элемента, который имеет наибольшую степень важности, исходя из разработанного алгоритма.

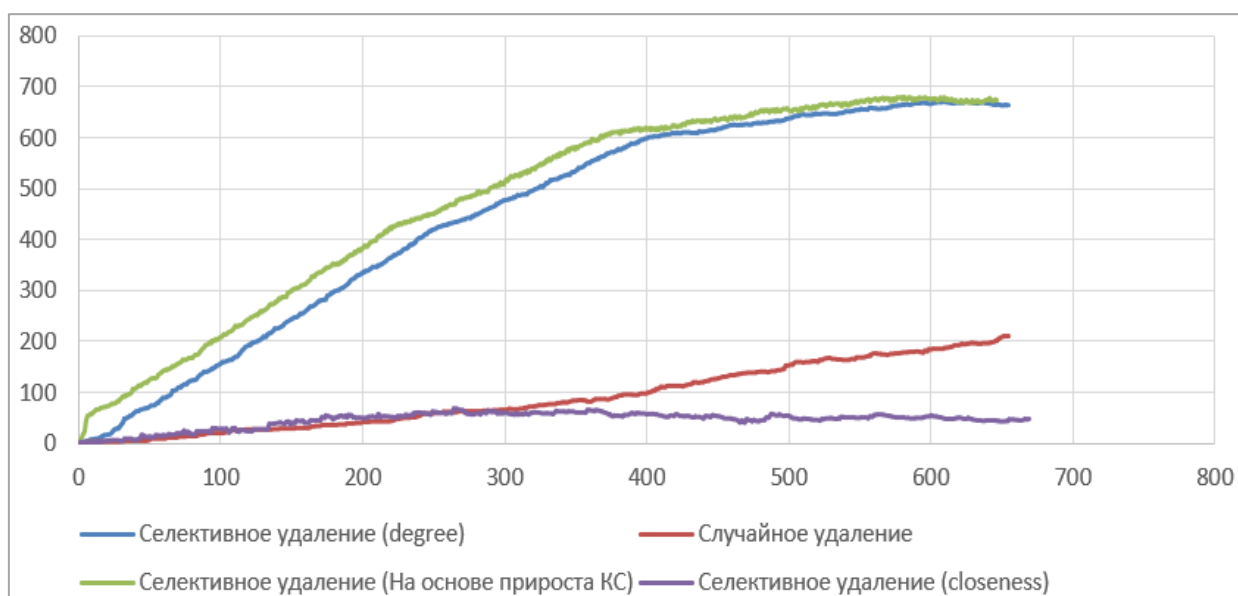


Рисунок 17 – Сравнение алгоритмов определения важности компонента электросети

График показывает зависимость Y – количества компонент связности от количества отключенных узлов – X , отсортированных в соответствии с той или иной метрикой. Надо заметить, что данный алгоритм является переборным. При современных расчётных мощностях такая сложность алгоритма является проблемой, только при очень больших объемах исходных данных. Вместе с тем более точная информация о свойствах электросети, может значительно снизить ущерб от случайного или умышленного её повреждения, что зачастую влечет большие экономические потери, в сравнении со стоимостью расчетов.

Было проведено тестирование на экспериментальных данных об авариях трансформаторов. После произведения классификации каждым, из вышеперечисленных методом, были рассчитаны точности по метрике F-мера.

Результаты классификации, с использованием метода логистической регрессии, представлены в таблицах метрик точности по группам и матрице ошибок, (таблицы 3, 4).

Таблица 3 – Точность по метрикам precision, recall, f1 по группам. Логистическая регрессия

Группа	precision	recall	f1-score	support
0	0.00	0.00	0.00	3
1	0.47	0.88	0.61	8
2	0.75	0.67	0.71	9
3	0.67	0.89	0.76	9
4	0.00	0.00	0.00	2
5	1.00	0.33	0.50	3
6	0.00	0.00	0.00	3

Таблица 4 – Матрица ошибок. Логистическая регрессия

0	3	0	0	0	0	0
0	7	1	0	0	0	0
0	2	6	1	0	0	0
0	0	1	8	0	0	0
0	0	0	2	0	0	0
0	1	0	0	0	1	1
0	2	0	1	0	0	0

Результаты классификации, с использованием байесовского классификатора, представлены в таблицах метрик точности по группам и матрице ошибок, (таблицы 5, 6).

Таблица 5 – Точность по метрикам precision, recall, f1 по группам. Байесовский классификатор

Группа	precision	recall	f1-score
0	0.67	0.67	0.67
1	0.64	0.88	0.74
2	0.83	0.56	0.67
3	0.88	0.78	0.82
4	0.00	0.00	0.00
5	0.75	1.00	0.86
6	0.00	0.00	0.00

Таблица 6 – Матрица ошибок. Байесовский классификатор

2	1	0	0	0	0	0
0	7	1	0	0	0	0
0	3	5	0	1	0	0
0	0	0	7	2	0	0
0	0	0	1	0	0	1
0	0	0	0	0	3	0
1	0	0	0	1	1	0

Результаты классификации, методом опорных векторов, представлены в таблицах метрик точности по группам и матрице ошибок, (таблицы 7, 8).

Таблица 7 – Точность по метрикам precision, recall, f1 по группам. Метод опорных векторов

Группа	precision	recall	f1-score
0	0.50	0.67	0.57
1	0.75	0.75	0.75
2	0.86	0.67	0.75
3	0.80	0.89	0.84
4	0.50	0.50	0.50
5	0.67	0.67	0.67
6	0.00	0.00	0.00

Таблица 8 – Матрица ошибок. Метод опорных векторов

2	0	0	0	0	1	0
0	6	1	1	0	0	0
0	2	6	1	0	0	0
0	0	0	8	0	0	1
0	0	0	0	1	0	1
0	0	0	0	0	2	1
2	0	0	0	1	0	0

Результаты классификации, с использованием дерева решений, представлены в таблицах метрик точности по группам и матрице ошибок, (таблицы 9, 10).

Таблица 9 – Точность по метрикам precision, recall, f1 по группам. Дерево решений

Группа	precision	recall	f1-score
0	0.50	0.33	0.40
1	0.67	0.75	0.71
2	0.86	0.67	0.75
3	0.75	0.67	0.71
4	0.00	0.00	0.00
5	0.50	1.00	0.67
6	0.00	0.00	0.00

Таблица 10 – Матрица ошибок. Дерево решений

1	1	0	0	0	1	0
0	6	1	1	0	0	0
0	2	6	1	0	0	0
0	0	0	6	1	0	2
0	0	0	0	0	1	1
0	0	0	0	0	3	0
1	0	0	0	1	1	0

Сводные результаты по каждому из использованных методов представлены в таблице 11.

Таблица 11 – Первичные результаты классификации

Метод	Точность (F-мера)
Логистическая регрессия	0,53
Наивный байесовский классификатор	0,65
Метод К-ближайших соседей	0,65
Метод опорных векторов	0,68
Дерево решений	0,59

Метод, показавший наилучшую точность (0,68) – Метод опорных векторов.

Был проведен отбор признаков. После удаления менее значимых признаков были рассчитаны точности прогнозирования методов, результаты представлены в таблице 12.

Таблица 12 – Результаты классификации после исключения части признаков

Метод	Точность (F-мера)
Логистическая регрессия	0,43
Наивный байесовский классификатор	0,54
Метод К-ближайших соседей	0,55
Метод опорных векторов	0,48
Дерево решений	0,52

Поскольку исключение признаков, снизило точность прогнозирования, для последующих шагов, использовалась первичная модель. После чего была проведена оптимизация каждого из методов.

Оптимизация повысила точность всех вышеперечисленных методов, однако, более высокую точность показали метод К-ближайших соседей и метод опорных векторов.

Для метод К-ближайших соседей, наилучшим параметром для является $K = 4$. После оптимизации точность прогнозирования методом К-ближайших соседей возросла до 0,70.

Кривая оптимизации для метода опорных векторов представлена на графике 18.



Рисунок 18 – Кривая оптимизации метода опорных векторов

Наилучшими параметрами для метода опорных векторов является $C=100$. После оптимизации точность прогнозирования методом опорных векторов возросла до 0,76.

Аналогичных стек методом, был применен к набору тестовых данных об авариях на линиях электропередач. Классификация сигналов напрямую на

дала результатов с приемлемой точностью. Также, за счет высокой плотности записи сигнала обработка данных требовала заметного вычислительного времени. В связи с этим, был предложен новый подход, основанный на классификации дисперсий сигналов, а не сигналов напрямую. Данный подход базируется на предположении о том, что сигналы, содержащие аварию, обладают большим уровнем шума, чем остальные записи.

Результат классификации дисперсий по группам представлены в таблице 13.

Таблица 13 – Результаты классификации дисперсий сигналов. Метод к-ближайших соседей

Группа	precision	recall	f1-score
0	0.94	1.00	0.97
1	0.76	0.06	0.10

Проведена оптимизация, наилучший метод – метод К-ближайших соседей, с параметром $K = 2$. Средняя точность полученного прогноза 0,52.

Можно заметить высокую точность идентификации группы 0 и низкую точность для группы 1. Такой эффект возникает по причине несбалансированности выборки. Решение данной проблемы – балансировка выборки, производилась путём уменьшения размеры группы 0.

Следующим этапом, позволившим улучшить точность прогнозирования, стала классификация параметров шума данных. Из исходных данных был извлечен шум. В качестве фильтра использовался медианный фильтр. Пример отфильтрованных данных зависимости величины сигнала I от времени t представлен на рисунке 19.

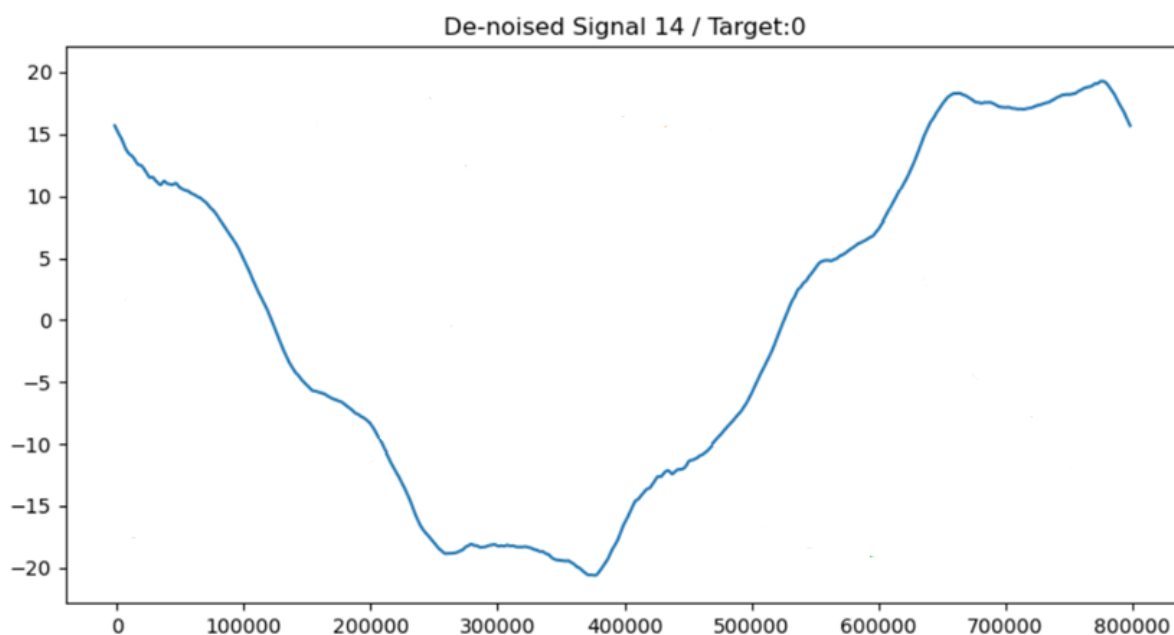


Рисунок 19 – Пример отфильтрованных данных

Вычитание отфильтрованного сигнала от исходного позволяет отделить шум. Пример графика зависимости величины шума от времени t представлен на рисунке 20.

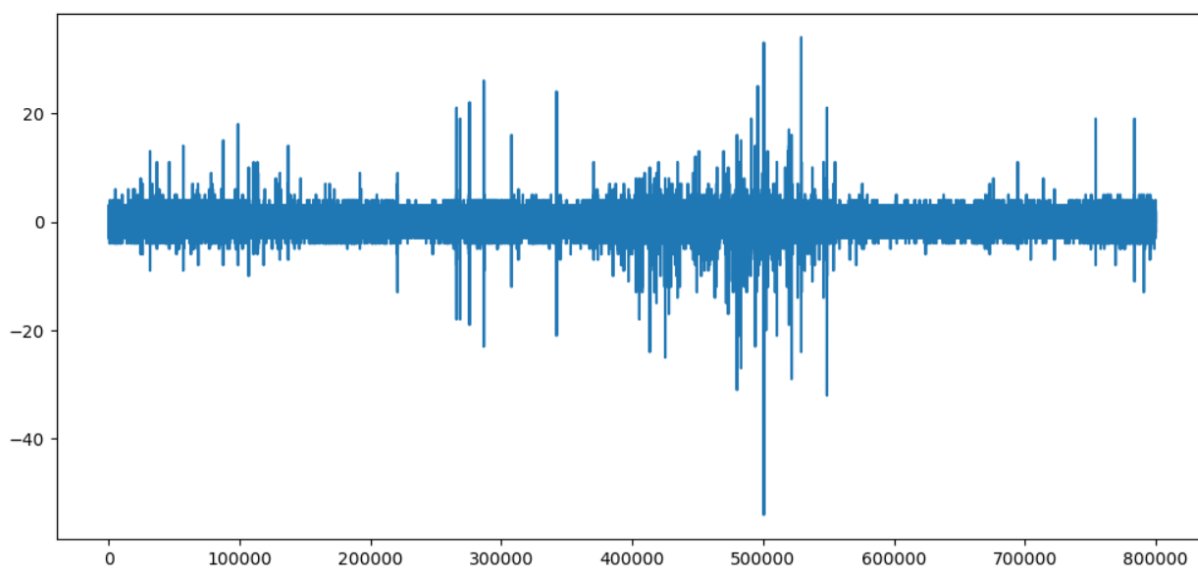


Рисунок 20 – График шума.

Наилучшим методом оказался метод К-ближайших соседей с параметром $K=13$.

Классификация дисперсий шума позволила повысить результаты прогнозирования до 0,63, таблица 14.

Таблица 14 – Результаты классификации дисперсий шума. Метод к-ближайших соседей

Группа	precision	recall	f1-score
0	0.70	0.79	0.74
1	0.58	0.46	0.51

Для измерения степени зашумленности данных рационально использовать не только дисперсии, но и другие дополнительные характеристики, определяющие наличия и степень шума.

В модель были добавлены следующие параметры:

1. минимальная величина шума;
2. максимальная величина шума;
3. среднее шума.

Результаты классификации модели с новыми параметрами представлены в таблице 15.

Таблица 15 – Результаты классификации параметров шума. Метод К-ближайших соседей

Группа	Precision	recall	f1-score
0	0.74	0.72	0.73
1	0.62	0.65	0.64

Средняя точность прогноза методом к-ближайших соседей 0,68.

Помимо, метки 0 и 1, показывающей наличие повреждения на определённой фазе, исходные данные содержат информацию о идентификатор группы сигналов. Было принято решение группировки аварий по фазам. Данная группировка позволит получать дополнительную информацию об аварии, поскольку теперь рассматриваемый сигнал будет состоять из трех записей, произведенных в один момент времени, для одного и того же прецедента аварии.

Одна точка полученной выборки содержит три группы следующих характеристик:

1. дисперсия;
2. минимальная величина шума;

3. максимальная величина шума;
4. среднее шума.

Пример зависимости величины сигнала I1, I2, I3 от времени t представлен на рисунке 21.

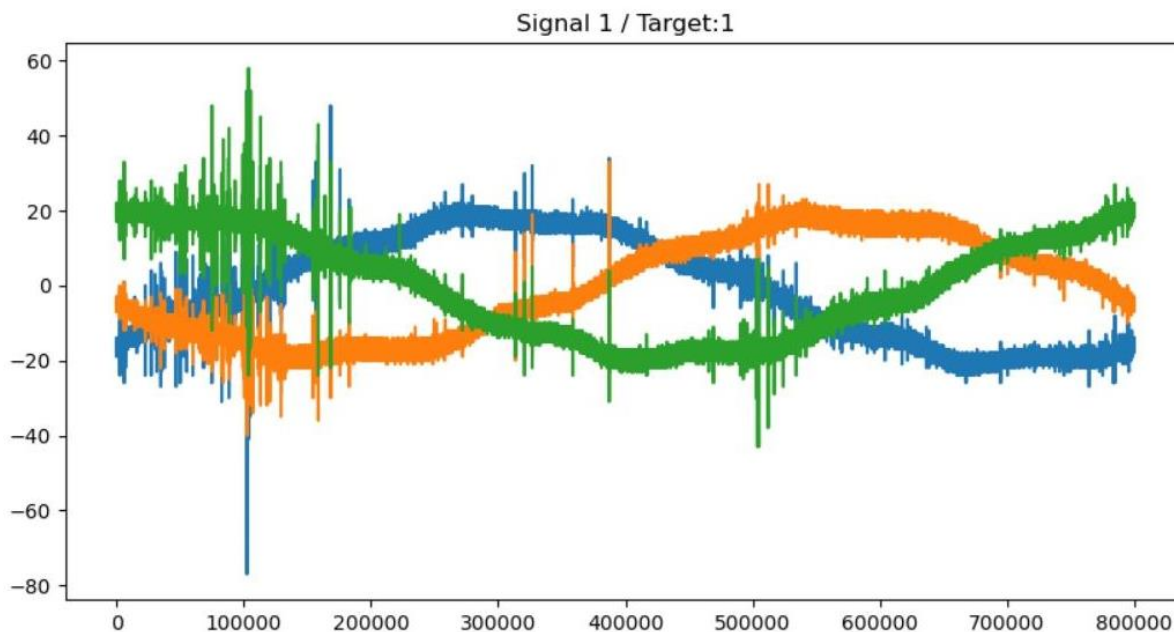


Рисунок 21 – Пример графика аварийной записи

Такой подход позволил заметно улучшить точность прогнозирования. После проведения оптимизации, наилучшим методом для прогнозирования стал метод к-ближайших соседей, с параметром $K = 5$. В таблице 16 представлены точности прогноза по группам.

Таблица 16 – Результаты классификации параметров шума с группами. Метод К-ближайших соседей.

Группа	precision	recall	f1-score
0	0.83	0.74	0.78
1	0.78	0.86	0.81

По метрике f-мера средняя точность прогноза 0,80. Для наиболее приоритетной группы 1, точность прогноза методом К-ближайших соседей составила 0,81.

Выводы

В вопросе прогнозирования аварий в электросетях, разработанных подход, на основе анализа параметров шума электрического сигнала, в совокупности с методами машинного обучения показал хорошую точность на экспериментальных данных. Это ещё раз подтверждает рациональность использование данного подхода в подобных задачах.

Разработанный алгоритм оценки важности электрокомпонента показала, в соответствии с выбранной метрикой, показал сравнимую эффективность с классическими метриками центральности узлов.

В целом применение разработанной системы позволит снизить влияние человеческого фактора на процесс детектирования и прогнозирования аварий в электросетях.

Повсеместное использование подобных систем, может значительно снизить расходы на эксплуатацию электросетей, с точки зрения поставщиков электроэнергии. А также, стабилизировать социально экономические процессы, подразумевающие непрерывную эксплуатацию электроэнергии, что позволит снизить экономические издержки потребителя, а также избежать проблем коммуникационного и технологического характера.

Заключение

Результаты работы

По результатам работы, был выполнен ряд задач.

На языке программирования Python, с использованием СУБД Influx и средства визуализации GRAFANA, был разработан модуль обработки временных данных от компонентов электросети.

При построении алгоритма обработки исходных данных, поступающих от датчиков системы, были реализованы следующие подзадачи:

1. Разработана функция чтения осциллограмм в формате COMTRADE и сохранения в базе данных временных рядов Influx.
2. Разработана функций визуализации осциллограмм.
3. Разработка функций анализа данных об аварийных событиях и технологических нарушениях в электрических системах.

Построение алгоритма оценки компонентов электросети, на основе прироста количества компонент связности и проведение сравнения данного алгоритма оценки с классическими метриками, осуществлено посредством реализации следующих подзадач:

1. Определены и продемонстрированы свойства сети. Рассчитаны связность графа и центральности вершин.
2. Построен алгоритм оценки важности компонента сети, с точки зрения последствий его отключения от сети.
3. Произведена оценка последствий селективного и случайного отключения компонентов, в соответствии с метриками closeness, degree и разработанной метрикой на основе прироста компоненты связности.

Разработка алгоритма прогнозирования вероятности отказа трансформатора на основе, данных анализа растворенных газов, с точностью результатов 76% и построение алгоритм прогнозирования вероятности аварии

на ЛЭП, на основе анализа шума, с точностью результатов 80%, осуществлено по средствам реализации следующих подзадач:

1. Произведен сбор информации о физических свойствах компонентов электросети.
2. Представлен обзор методов анализа данных об аварийных событиях и технологических нарушениях в электрических системах.
3. Произведено обоснование выбора метода прогнозирования.
4. Реализован алгоритм обработки исходных данных.
5. Реализован алгоритм прогнозирования на основе методов машинного обучения.

Разработан модуль визуализации электрокомпонентов и их параметров на географической карте.

Перспективы развития

Разработанная система имеет широкие возможности развития по нескольким направлениям.

По части графовых методов, является рациональным рассмотрение новых алгоритмов центральности узлов, а также использования альтернативных электротехнических или экономических метрик сравнения алгоритмов. Применение экономических метрик сравнения алгоритмов расчета центральностей позволит провести оптимизацию издержек, связанных с умышленным или случайным отключением компонента системы.

С точки зрения развития базы методов машинного обучения, использованных в системе, имеет смысл рассмотреть методы на основе нейронных сетей, а также ансамблевые методы. Однако применение большого количества новых методов, с одной стороны, позволит увеличить точность прогнозирования, с другой же стороны, увеличит расчетное время работы алгоритмов. Отсюда можно сделать вывод, что увеличение базы рассматриваемых методов машинного обучения рационально для крупных

систем, имеющих большие расчетные мощности, и обеспечивающих важные электросети регионального или федерального масштаба.

По части структуры разработанного программного обеспечения оптимизация архитектуры системы может рассматриваться в двух направлениях. Использование дополнительных библиотек и модулей, для визуализации аварий, или, наоборот, отказ от использования стороннего программного обеспечения и разработка крупномасштабной системы, с собственным географическим модулем и модулем визуализации. Первый тип преобразования системы подойдет для процесса разработки, при котором заказчик не желает брать риски и финансовую ответственность, связанные с поддержкой и обновлением программного обеспечения. Второй тип, подойдет для крупных систем, в которых есть кадровые ресурсы, для осуществления доработки и поддержки, но при этом необходима гарантия сохранности информации внутри системы.

Любой сценарий развития разработанной системы из предложенных приведет к качественному ее улучшению, однако для успешной реализации необходимо учитывать соответствующие риски и понимать все технические, экономические и социальные аспекты такого преобразования.

Список литературы

1. Алексеев В.Е., Таланов В.А. Графы. Модели вычислений. Структуры данных: Учебник //Нижний Новгород: Изд-во ННГУ. –2005. –307 с.
2. ГОСТ 19431. Энергетика и электрификация. Термины и определения
3. ГОСТ 24291. Электрическая часть электростанции и электрической сети. Термины и определения.
4. ГОСТ Р 55438-2013 Единая энергетическая система и изолированно работающие энергосистемы. Оперативно-диспетчерское управление. Релейная защита и автоматика. Взаимодействие субъектов электроэнергетики и потребителей электрической энергии при создании (модернизации) и эксплуатации. Общие требования (с Изменением N 1)
5. ГОСТ 16110-82 Трансформаторы силовые. Термины и определения
6. Ahmed Abu-Siada. Power Transformer Condition Monitoring and Diagnosis// The Institution of Engineering and Technology. – 2018. – 326 с.
7. Aida Asadi Majd, Haidar Samet & Teymoor Ghanbari. k-NN based fault detection and classification methods for power transmission systems//Prot Control Mod Power Syst 2. – 2017. – 32 с.
8. A. Reciou, B. Benseghier and H. Khalfallah. Power system fault detection, classification and location using the K-Nearest Neighbors//4th International Conference on Electrical Engineering (ICEE), Boumerdes. – 2015.– 287 с.
9. Bowen Kan, Wendong Zhu, Guangyi Liu, Xi Chen, Di Shi, Weiqing Yu. Topology Modeling and Analysis of a Power Grid Network Using a Graph Database// International Journal of Computational Intelligence Systems, Vol. 10. – 2017. – 1355–1363 с.
10. Božić Miloš, Stojanović Miloš, Stajić Zoran, Vukić D. Power transformer fault diagnosis based on dissolved gas analysis with logistic regression//Przeglad Elektrotechniczny – 2013. – 83-87 с.
11. C37.111-1999 - IEEE Standard Common Format for Transient Data Exchange (COMTRADE) for Power Systems.

12. C. Schneider, J. Staninovski, L. Cheim, J. Vines, S. Varad. Transformer Reliability Taking Predictive Maintenance Program to the Next Level// Cigre Study Committee A2. – 2017. – 10 c.
13. Debani Prasad Mishra, Papia Ray. Fault detection. Location and classification of a transmission line// Neural Comput & Applic. – 2018. – 1424 c.
14. Hui Hwang Goh, Sy yi, Sim, Asad Shaykh, Md.Humayun Kabir, Chin Wan Ling, Qing Shi Chua, Kai Chen Goh. Transmission Line Fault Detection: A Review// Indonesian Journal of Electrical Engineering and Computer Science. – 2017. – 205 c.
15. Fang Yuan, Jiang Guo, Zhihuai Xiao, Bing Zeng, Wenqiang Zhu, Sixu Huang. A Transformer Fault Diagnosis Model Based on Chemical Reaction Optimization and Twin Support Vector Machine// Energies. – 2019. – 18 c.
16. Matthew O. Jackson. Social and Economic Networks // Princeton University Press. – 2008. – 647 c.
17. M. Akbari, P. Khazaei. Failure Modes and Effects Analysis (FMEA) for Power Transformers// 28th International Power System Conference. – 2013. – 7 c.
18. Pallav Kumar Bera, Rajesh Kumar, Can Isik. Identification of Internal Faults in Indirect Symmetrical Phase Shift Transformers Using Ensemble Learning//arxiv.org – 1991. [Электронный ресурс]. URL: <https://arxiv.org/abs/1811.04537>
19. Prasenjit Shil, Ameren, Tom Anderson, SAS Institute Inc. Distribution Transformer Health Monitoring and Predictive Asset Maintenance// SAS global forum. – 2019. – 3944 c.
20. Raji Murugan, Raju Ramasamy. Failure analysis of power transformers for effective maintenance planning in electric utilities//Engineering Failure Analysis. – 2015. – 182-192 c.
21. Ricard V. Sole, Marti Rosas-Casals, Bernat Corominas-Murtra and Sergi Valverde. Robustness of the European power grids under intentional attack

//arxiv.org. – 1991. [Электронный ресурс]. URL:
<https://arxiv.org/abs/0711.3710v1>

22. Sayari Das, Shiv P. Sing, Bijaya K. Panigrahi. Transmission line fault detection and location using Wide Area Measurements// Electric Power Systems Research– 2017. – 96-105 с.

23. S. Souahlia, K. Bacha and A. Chaari. SVM-based decision for power transformers fault diagnosis using Rogers and Doernenburg ratios DGA//10th International Multi-Conferences on Systems, Signals & Devices. – 2013. – 495 с.

24. Valabhoju Ashok, Anamika Yadav. A real-time fault detection and classification algorithm for transmission line faults based on MODWT during power swing//Wiley Online Library. – 2019. [Электронный ресурс]. URL:
<https://onlinelibrary.wiley.com/doi/full/10.1002/2050-7038.12164>

25. Vandermaar John. Condition Assessmene of Power Transformers// Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century. – 2008. – 168 с.

26. Zhang Zhe-wen, Wang Yong, Ying Ding, Tian lei, Zhou Ying-jian. A Transformer Fault Diagnosis Method Based on Bayesian Network// Journal of Physics: Conference Series. – 2019. – 8 с.